outlines as line art does. Even where images possess clear outlines, such as for a black building against a blue sky, the edges are poorly aliased and differ significantly from computer generated images. In particular embodiments, higher compression ratios may indicate a lower probability of the image being legitimate. Natural photographs tend to compress poorly, whereas graphics with large unbroken regions of a solid color compress well. In particular embodiments, the aspect ratio may indicate whether an image is a legitimate natively-captured image. For example, cameras tend to have standard aspect ratios, such as 4:3 or 16:9. Images generated by applications may have different non-camera aspect ratios. Images having such ratios are far more likely to be illegitimate rather than natively captured.

[0029] In particular embodiments, photo spam detector **260** may also extract metadata, such as exchangeable image file format (EXIF) data from the image itself. This EXIF data is of little use in and of itself, because it is a relatively elementary task to fake EXIF data. However, the concurrence between the EXIF data and the image properties may be useful in determining the legitimacy of the image. In particular embodiments, discrepancies between the extracted image features and the EXIF data may indicate that the image is illegitimate. For example, if the aspect ratio or pixel dimensions indicated by the EXIF data differs from the actual uploaded image, the image is more likely to be illegitimate. Methods of verifying consistency between the EXIF data and image data are not limited to dimensions. In particular embodiments, specialized image feature extraction modules may extract particular features to verify against certain metadata. For example, photos show with low apertures, such as F/2 and below, should be blurry in one area and focused in another. Edge detection algorithms may be used to detect increased areas of aliasing to verify that one area is sharp and others are blurry. Conversely, photos taken with a high aperture, such as F/10 and above, should be essentially completely in focus. Increased aliasing in one area may indicate inconsistency between the EXIF aperture data and the actual image. This disclosure contemplates any suitable algorithm or methodology for verifying concurrence between metadata and image characteristics.

[0030] Features extracted by image feature extraction module **301** are fed into probabilistic model **302**. Probabilistic model **302** receives the extracted image features as inputs and, from those inputs, calculates a probability as to whether the image from which the features were extracted is a legitimate natively-captured image or an illegitimate image. Probabilistic model **302** may be any suitable type of machine learning application. In particular embodiments, probabilistic model **302** comprises a neural network. In particular embodiments, probabilistic model **302** is a support vector regression (SVR). When utilizing an SVR, an initial training set of any number of legitimate and illegitimate images may be fed into the machine, from which the machine learns what extracted features are more likely to be legitimate images or illegitimate images. In particular embodiments, the probabilistic model is a nonlinear classifier. In particular embodiments, the SVR may be multiclass. In particular embodiments, the training of the probabilistic model may be supervised or unsupervised. In particular embodiments, probabilistic model **302** may be subject to feedback from users. For example, users may flag images as illegitimate. When such an event occurs, it is added to the training set and probabilistic model **302** is updated. In particular embodiments, the probabilistic model creates one or more probability density functions (PDFs) that determine the probability an image is illegitimate for a given input feature. This disclosure contemplates any suitable machine learning algorithm or application for implementing probabilistic model **302**.

[0031] Photo spam detector **260** in particular embodiments also includes OCR (optical character recognition) module **305**. Illegitimate images often contain large blocks of text. In particular embodiments, OCR module **305** may search the image in question for large blocks of text, and flag the image as less likely to be legitimate based upon this detection. While the existence of text blocks in images is not dispositive of its legitimacy; for example, a legitimate natively-captured image may include captured text contained in real-life signage. However, in particular embodiments, the disparity of the signage from the rest of the image; i.e., sharply aliased, significantly different gamma value, placement, etc. may indicate that the image is illegitimate. This disclosure contemplates any manner of combining other extracted image features with optical character recognition to calculate the probability that an image containing blocks of text is illegitimate.

[0032] Photo spam detector **260** also includes tag analyzer module **303**. Tag analyzer module may process any set of requests to tag a photo. In particular embodiments, tag analyzer module is run concurrently with image feature extraction module **301**. In particular embodiments, it is run separately from image feature extraction module **301**. Tag analyzer module **301** may access accounts module **220**. In particular embodiments, tag analyzer module monitors the temporal rate that the tag requests are received. For example, if the tags are received substantially simultaneously, or under a predetermined threshold from each other, such as 3/10 of a second, it s likely that the tags are auto-generated by an application and constitute spam. In particular embodiments, tag analyzer module **301** analyzes the relationship between the users associated with the tags to determine the likelihood that the tags are spam. For example, if the tag requests are received in alphabetical order by username (either first or last), there is a high likelihood that the tags were automatically generated by an application or script, and constitute spam. In other embodiments, tag analyzer module **303** accesses social relationship information about the users associated with the tag requests from accounts module **220**. Tag analyzer module may take into account coefficient scores between the set of users tagged. Coefficient scores may be calculated based on any number of variables. For example, increased interaction between any two users may result in a higher coefficient score, lending to the inference that the two are closer real-world friends. Similarly, a large number of mutual friends may increase the coefficient score. Methods for calculating coefficient scores between two or more users are well-known. Regardless of how the coefficient scores are calculated, once tag analyzer module **303** obtains coefficient scores between each of the set of tagged users, tag analyzer **303** may calculate the probability that the image is legitimate based on the coefficient scores. For example, a set of users with coefficient scores below a predetermined threshold, indicating that the users are not very close friends in real-life, are less likely to be tagged in the same photo. Thus the probability the tags are spam is increased.